

## SARS

Matthews, Thomas; Triantis, Kostas; Whittaker, Robert J.; Guilhaumon, Francois

DOI:

[10.1111/ecog.04271](https://doi.org/10.1111/ecog.04271)

License:

Other (please specify with Rights Statement)

*Document Version*

Peer reviewed version

*Citation for published version (Harvard):*

Matthews, T, Triantis, K, Whittaker, RJ & Guilhaumon, F 2019, 'SARS: an R package for fitting, evaluating and comparing species–area relationship models', *Ecography*, vol. 42, no. 8, pp. 1353-1457.  
<https://doi.org/10.1111/ecog.04271>

[Link to publication on Research at Birmingham portal](#)

### **Publisher Rights Statement:**

Checked for eligibility: 19/03/2019

This is the peer reviewed version of the following article: Matthews, T. J., Triantis, K. A., Whittaker, R. J. and Guilhaumon, F. (2019), sars: an R package for fitting, evaluating and comparing species–area relationship models. *Ecography*. doi:10.1111/ecog.04271., which has been published in final form at: <https://doi.org/10.1111/ecog.04271>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

### **General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### **Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# sars: an R package for fitting, evaluating and comparing species–area relationship models

Thomas J. Matthews<sup>1,2</sup>, Kostas Triantis<sup>3</sup>, Robert J. Whittaker<sup>4,5</sup> & François Guilhaumon<sup>6</sup>

<sup>1</sup>GEES (School of Geography, Earth and Environmental Sciences), and Birmingham Institute of Forest Research, The University of Birmingham, Birmingham, B15 2TT

<sup>2</sup>CE3C – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group and Univ. dos Açores – Depto de Ciências e Engenharia do Ambiente, PT-9700-042, Angra do Heroísmo, Açores, Portugal

<sup>3</sup>Department of Ecology and Taxonomy, Faculty of Biology, National and Kapodistrian University of Athens, Athens GR-15784, Greece

<sup>4</sup>School of Geography and the Environment, University of Oxford, South Parks Road, Oxford, OX1 3QY, UK

<sup>5</sup>Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

<sup>6</sup>MARBEC, IRD, Ifremer, Univ Montpellier, CNRS, Montpellier, France

**Corresponding author:** Thomas J. Matthews, School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK. E-mail: t.j.matthews@bham.ac.uk

**Decision date:** 11-Mar-2019

---

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1111/ecog.04271].

## Abstract

The species–area relationship (SAR) constitutes one of the most general ecological patterns globally. A number of different SAR models have been proposed. Recent work has shown that no single model universally provides the best fit to empirical SAR datasets: multiple models may be of practical and theoretical interest. However, there are no software packages available that a) allow users to fit the full range of published SAR models, or b) provide functions to undertake a range of additional SAR-related analyses. To address these needs, we have developed the R package ‘sars’ that provides a wide variety of SAR-related functionality. The package provides functions to: a) fit 20 SAR models using non-linear and linear regression, b) calculate multi-model averaged curves using various information criteria, and c) generate confidence intervals using bootstrapping. Plotting functions allow users to depict and scrutinize the fits of individual models and multi-model averaged curves. The package also provides additional SAR functionality, including functions to fit, plot and evaluate the random placement model using a species–sites abundance matrix, and to fit the general dynamic model of oceanic island biogeography. The ‘sars’ R package will aid future SAR research by providing a comprehensive set of simple to use tools that enable in-depth exploration of SARs and SAR-related patterns. The package has been designed to allow other researchers to add new functions and models in the future and thus the package represents a resource for future SAR work that can be built on and expanded by workers in the field.

## Keywords

Diversity–area relationship, general dynamic model, island biogeography, multi-model inference, power model, random placement model, sars, species accumulation curve, species–area relationship

## BACKGROUND

The species–area relationship (SAR) describes the near universally observed pattern whereby the number of species increases with the area sampled, and it has been described as one of ecology’s few laws (Rosenzweig 1995). The SAR is a fundamental component of numerous ecological and biogeographical theories, such as the equilibrium theory of island biogeography (MacArthur and Wilson 1967). In addition, SAR models have been widely used in applied ecology and conservation biogeography: for example, to predict the number of extinctions due to habitat loss (May et al. 1995). Numerous types of SAR have been described, and one primary dichotomy employed is the split of SARs into island SARs (ISARs), whereby each data point is an individual island or isolated sample, and species accumulation curves (SACs) that represent cumulative counts of increased species number with sampling area (Gray et al. 2004; see Scheiner 2003 for an alternative SAR typology). Whilst the remainder of the paper and the described R package are focused on ISARs, the models and the model fitting procedure can equally be applied to SACs (see Matthews et al. 2016a), although it should be noted that in SACs the data points are not independent of one another.

Over 20 SAR models have been described in the literature (Dengler 2009, Tjørve 2009, Triantis et al. 2012). However, despite this wide range of models, the majority of SAR studies are still based exclusively on the power model (Arrhenius 1921), which if fitted in its non-linear (untransformed) form generally takes a convex form. Often, the log–log representation of the power model is used as it can be fitted using standard linear regression, and its parameters are more easily interpretable (Rosenzweig 1995). However, whilst the power model has been found to provide a reasonable fit to a wide range of datasets (Dengler 2009, Triantis et al. 2012, Matthews et al. 2016b), it is not

universally the best model, and a number of studies have reported other models to provide better fits to empirical data (e.g. Triantis et al. 2012, Benchimol and Peres 2013, Matthews et al. 2016b). The possibility of scale dependency of the form of the SAR has long been of interest, with, for example, a theoretical case being made for SARs at intermediate spatial scales being approximated by a power model, whilst at larger spatial scales the form of the SAR has been theorised to be sigmoidal (He and Legendre 1996, Lomolino 2000). Additionally, it is only recently that the SAR for archipelagos as units of analysis and not just islands has started to be studied (e.g. Triantis et al. 2015), and thus we know little about the form of archipelago SARs.

Due to the increased recognition of model uncertainty in SAR research, a number of recent studies have employed a multi-model inference approach (Burnham and Anderson 2002) in the analysis of SARs, whereby either (1) multiple SAR models are compared using various criteria (e.g. AIC) and a best model is chosen (e.g. Dengler 2009, Benchimol and Peres 2013), or (2) multiple SAR models are fitted and a multi-model averaged curve is calculated using, for example, AIC weights (e.g. Guilhaumon et al. 2008). We are not aware of any published software package that enables users to fit, and create multi-model averaged curves using more than eight SAR models.

Considering currently available software, the BAT R package (Cardoso et al. 2014) provides functions to fit three SAR models (linear, power and logarithmic); however, this package is focused on general biodiversity assessment and thus does not provide any additional SAR functionality. The mmSAR R package (Guilhaumon et al. 2010) is focussed on SARs and while it allows users to fit eight SAR models using an information theoretic framework, it does not include several models that have been

found to provide the best fits to several empirical datasets (Triantis et al. 2012, Matthews et al. 2016b).

To provide a set of tools to fill these gaps, we have developed the R package ‘sars’. The package provides functions to fit 20 SAR models using non-linear and linear regression, calculate multi-model averaged curves using various information criteria, and generate confidence intervals using bootstrapping. Novel features compared with mmSAR include (i) user-friendly functions for plotting (the user can now plot weighted multimodel SAR curves along with the individual SAR model curves) and (ii) determining the observed shape of the model fit (i.e. linear, convex up, convex down or sigmoidal) and (iii) presence or not of an asymptote, and (iv) functions to fit, plot and evaluate Coleman’s (1981) random placement model using a species-site abundance matrix, and (v) to fit the general dynamic model (GDM) of island biogeography (Whittaker et al. 2008). In addition, the mmSAR package (which has been deprecated) no longer complies with recognised programming good practice (Wickham 2015), is not on CRAN (the main repository of R packages), and is not user friendly (e.g. it requires the user to load individual models prior to fitting). There was therefore a need to design a new package from scratch.

## **METHODS AND FEATURES**

The ‘sars’ (species–area relationships) package has been programmed using standard S3 methods and is available on CRAN (version 1.1.1; R Core Team 2017), with the development version on GitHub (txm676/sars), meaning researchers can easily add in their own models and functions and integrate these into the multi-model inference

framework. Thus, the package represents a resource for future SAR work that can be built on and expanded by workers in the field.

### **Fitting individual SAR models and a set of SAR models**

The package provides functions to fit each of the 20 SAR models listed in Table 1 (see also Triantis et al. 2012). The ‘sar\_models’ function can be used to bring up a list of the 20 model names and Table 1 can be generated in the package using the ‘display\_sars\_models’ function. With the exception of the linear model (which is fitted using standard linear regression), all models are fitted using non-linear regression and the model parameters are estimated by minimizing the residual sum of squares with an unconstrained Nelder-Mead optimization algorithm and the ‘optim’ R function (R Core Team 2017). The starting values for the parameter estimates are carefully chosen to avoid numerical problems and to speed up the convergence process. However, custom starting values can be provided for any of the 20 models using the ‘start’ argument in the model fit functions.

Each individual model fit returns an object of class ‘sars’, which is a list of 22 elements containing relevant model fit information, such as the model parameter estimates, the fitted values, the residuals, model fit statistics (e.g. AIC,  $R^2$ ), the observed model shape (linear, convex or sigmoidal), whether or not the fit is asymptotic, and convergence information. The returned object can easily be plotted using the ‘plot.sars’ generic function; as this function is based on the base R plotting framework, the plot aesthetics can be edited using standard plotting arguments (see the ‘plot.sars’ documentation in the package). Summary and print generic functions are also provided for class ‘sars’; these

functions follow the output of the standard 'lm' function in the 'stats' R package (R Core Team 2017).

Multiple SAR models can be fitted to the same dataset using the 'sar\_multi' function and the resultant  $n$  model fit objects stored together as a 'fit\_collection' object. This object is a list of class 'sars', where each of the  $n$  elements contains an individual SAR model fit. Using the 'plot.sars' generic function on a 'fit\_collection' object generates a grid of the  $n$  individual model fit plots (Fig. 1). The output from all of the examples given in the paper can be found in the package vignette (available from CRAN and GitHub).

*#load an example dataset (Preston, 1962), fit the logarithmic SAR model, #return a model fit summary and plot the model fit.*

```
library(sars)
data(galap)
fit <- sar_loga(data = galap)
summary(fit)

##
## Model:
## Logarithmic
##
## Call:
## S == c + z * log(A)
##
## Did the model converge: TRUE
##
## Residuals:
##      0%      25%      50%      75%     100%
## -164.000  -25.675   11.350   38.725  115.700
##
## Parameters:
##      Estimate Std. Error    t value    Pr(>|t|)      2.5%
## 97.5%
## c   0.2915341   34.5985767   0.0084262   0.9933958  -68.9056193
## 69.489
## z  30.2802630    8.3203931   3.6392827   0.0026812  13.6394768
## 46.921
##
## R-squared: 0.49, Adjusted R-squared: 0.41
## AIC: 141.78, AICc: 143.78, BIC: 144.1
## Observed shape: convex up, Asymptote: FALSE
```



```
##
##
## Warning: The fitted values of the model contain negative values
(i.e. negative species richness values)

plot(fit)

#Create a fit_collection object containing multiple SAR model fits,
and #plot all fits.

fitC <- sar_multi(data = galap, obj = c("power", "loga", "monod"))

##
## Now attempting to fit the 3 SAR models:
##
## -- multi_sars -----
multi-model SAR --
## > power : v
## > loga : v
## > monod : v

plot(fitC) #see Fig.1
```

## Model fit validation

Model fits can be evaluated through tests of the normality and homoscedasticity of the residuals. Any of three tests can be selected to test the normality of the residuals: 1) the Lilliefors extension of the Kolmogorov normality test (the default; `normaTest = "lillie"`), 2) the Shapiro-Wilk test of normality (to be preferred when sample size is small; `"shapiro"`), and 3) the Kolmogorov-Smirnov test (`"kolmo"`). Alternatively, an option to omit a residuals normality test is provided (`"none"`). Three options are provided to check for the homogeneity of the residuals: 1) a correlation of the residuals with the model fitted values (the default; `"cor.fitted"`), 2) a correlation of the residuals with the area values (`"cor.area"`), or 3) no homogeneity test (`"none"`). If a test is selected and is significant at the 5% level a warning is provided in the model summary; alternatively, the full results of the three tests can be accessed in the model fit output. A third model validation check for negative predicted richness values (i.e. when at least

one of the fitted values is negative) is automatically undertaken and a warning is provided in the model summary if negative values are predicted.

*#load an example dataset, fit the linear SAR model whilst running residual #normality and homogeneity tests, and return the results of the residual #normality test*

```
data(galap)
```

```
fit <- sar_linear(data = galap, normaTest = "lillie", homoTest =  
  "cor.fitted")
```

```
summary(fit) #a warning is provided indicating the normality test  
failed
```

```
##  
## Model:  
## Linear model  
##  
## Call:  
## S == c + m*A  
##  
## Did the model converge: TRUE  
##  
## Residuals:  
##      0%      25%      50%      75%     100%  
## -107.90  -52.10  -24.15   31.75  218.00  
##  
## Parameters:  
##      Estimate Std. Error  t value  Pr(>|t|)    2.5 %   97.5 %  
## c 70.314003   25.743130   2.731370  0.016228 15.100481 125.5275  
## m  0.185382    0.072884   2.543504  0.023410  0.029060  0.3417  
##  
## R-squared: 0.32, Adjusted R-squared: 0.21  
## AIC: 146.36, AICc: 148.36, BIC: 148.68  
## Observed shape: linear, Asymptote: FALSE  
##  
##  
## Warning: The normality test selected indicated the model residuals  
are not normally distributed (i.e. P < 0.05)
```

```
fit$normaTest
```

```
## $test  
## [1] "lillie"  
##  
## [[2]]  
##  
##  Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data:  res  
## D = 0.2146, p-value = 0.04725
```

### **Observed model shape and identifying an asymptote**

Whilst each of the 20 models has a general shape (Tjørve 2009, Triantis et al. 2012; see Table 1), the actual observed shape of the model fit can be different, for some models, depending on the parameter estimates. This is important as the shape of the curve has significant implications for conservation applications and the testing of macroecological theory (Rosenzweig 1995, Guilhaumon et al. 2008, Triantis et al. 2012). In ‘sars’, the observed shape of a model fit is determined using the sequential algorithm outlined in Triantis et al. (2012). The shape is calculated using the model fit within the observed range of area values. Briefly, the algorithm works by first determining whether the fit is a straight line. Then, if the fit is classified as not being linear, the observed shape is classified as either convex or sigmoidal by analysis of the second derivative (with respect to area) of the model fit (the full algorithm is detailed in Triantis et al. 2012, p. 220). There has also been considerable debate in the SAR literature as to whether or not the SAR is asymptotic (e.g. Williamson et al. 2001). In the ‘sars’ package, to determine whether a fit is asymptotic, for the relevant models the fitted model parameters are analysed to check whether the estimated asymptote is within the range of the sample data (Triantis et al. 2012).

### **Multimodel SAR curve**

As well as fitting individual models, the package provides a function (‘sar\_average’) to fit up to 20 models, compare the resultant fits using information criteria, and construct a multimodel-averaged SAR curve based on information criteria weight (see Burnham and Anderson 2002, Guilhaumon et al. 2008, 2010). The multimodel average curve is constructed as a linear combination of individual model fits by multiplying the

predicted richness values of each of the successfully fitted models by the model's information criterion weight, and then summing the resultant values across all models (Burnham and Anderson 2002). Three information criteria are available in the package: AIC, AIC<sub>c</sub> and BIC (see Burnham and Anderson 2002). Confidence intervals around the multimodel averaged curve can be calculated using a non-parametric bootstrap algorithm described in Guilhaumon et al. (2010). Briefly, each of the SAR models used in the 'sar\_average' function is fitted to the data, and the fitted values and residuals stored. The residuals are then transformed using the approach in Davison and Hinkley (1997, p.259). For each bootstrap sample, an individual model fit is selected with the probability of selection being equal to that model's information criterion weight. The transformed residuals from this fit are then sampled with replacement and added to the model's fitted richness values. The 'sar\_average' function is then used to fit all candidate SAR models to this bootstrapped set of response values, and the multimodel averaged fitted values stored. Percentile confidence intervals are then calculated using all bootstrapped fitted values, following Buckland et al. (1997).

The 'sar\_average' function can be used without specifying any models, in which case the package attempts to fit each of the 20 models in Table 1; alternatively, a vector of model names or a 'fit\_collection' object (generated using the 'sar\_multi' function) can be provided using the 'obj' argument. The three model validation tests listed above (normality and homogeneity of residuals, and negative predicted values) can be selected; if any model fails one or more of the tests during the fitting process it is removed from the resultant multimodel SAR curve. The output of the 'sar\_average' function is a list of class 'multi' and class 'sars', with two elements. The first element ('mmi') contains the multi model inference (fitted values of the multimodel SAR

Accepted Article

curve), and the second element ('details') contains a range of information regarding the fitting process, including the successfully fitted models, the models removed due to failing any of the validation tests, and the information criterion values, delta values and weights for the successfully fitted models. The returned object can easily be plotted using the 'plot.multi' generic function, and multiple plot options are available (using the 'type' argument; see Fig. 2). The fits of all the successfully fitted models and the multimodel SAR curve (with or without confidence intervals) can be plotted together ('type' = "multi"), and a barplot of the information criterion weights of each model can also be produced ('type' = "bar").

*#load an example dataset (Niering, 1963), run the 'sar\_average' function #using a vector of model names and with no model validation tests, and #produce the plots in Figure 2 of the paper*

```
data(niering)

#run the 'sar_average' function using a vector of model names
fit <- sar_average(data = niering, obj =
  c("power","loga","koba","mmf","monod","negexpo","chapman",
    "weibull3","asyp"), normaTest = "none", homoTest =
    "none", neg_check = FALSE, confInt = TRUE, ciN = 50)

##
## Now attempting to fit the 9 SAR models:
##
## -- multi_sars -----
## multi-model SAR --
## > power      : v
## > loga       : v
## > koba       : v
## > mmf        : v
## > monod      : v
## > negexpo    : v
## > chapman    : Warning: could not compute parameters statistics
## > weibull3   : v
## > asymp      : x
##
## 1 models could not be fitted and have been excluded from the multi
SAR
##
## All models passed the model validation checks
##
## 8 remaining models used to construct the multi SAR:
```

```

## Power, Logarithmic, Kobayashi, MMF, Monod, Negative exponential,
Chapman Richards, Cumulative Weibull 3 par.
## -----
##
## Calculating sar_multi confidence intervals - this may take some
time:

par(mfrow = c(3,1))

#plot all model fits with the multimodel SAR curve
plot(fit, ModTitle = "a) Multimodel SAR")

#plot the multimodel SAR curve (with confidence intervals; see
explanation #in the main text, above) on its own
plot(fit, allCurves = FALSE, ModTitle = "c) Multimodel SAR with
confidence intervals", confInt = TRUE)

#Barplot of the information criterion weights of each model
plot(fit, type = "bar", ModTitle = "b) Model weights", cex.lab = 1.3)

```

### Additional functions

In addition to the main functions used to fit and compare the 20 SAR models, the ‘sars’ package provides additional functions for specific SAR-based analyses. First, a function is provided to fit the log–log version of the power model (a function that is often fitted in SAR studies; Rosenzweig 1995) and compare parameter values with those generated using the non-linear power model. The log–log version of the power model is not equivalent to its non-linear counterpart because of non-equivalence in the study of the variation in a variable and in its transformation (Gitay et al. 1991, He and Legendre 1996), and bias of back-transformed results obtained on a logarithmic scale (Fattorini 2007). Second, a function has been added that enables the fitting of Coleman's (1981) random placement model to a species/sites abundance matrix. According to this model, the number of species occurring on an island depends on the relative area of the island and the regional relative species abundances. The fit of the random placement model

can be determined through use of a diagnostic plot (which can be generated using the ‘plot.coleman’ function) of island area (log transformed) against species richness, alongside the model’s predicted values (see Wang et al. 2010, Matthews et al. 2015). Following Wang et al. (2010), the model is rejected if more than a third of the observed data points fall beyond one standard deviation from the expected curve. Finally, a function is provided to fit the general dynamic model of island biogeography (Whittaker et al. 2008) using three different SAR models (linear, logarithmic and power).

*#load an example dataset, fit the log-log power model, return a model fit #summary and plot the model fit. When ‘compare’ == TRUE, the non-linear #power model is also fitted and the resultant parameter values compared. #If any islands have zero species, a constant (‘con’) is added to all #species richness values.*

```
data(galap)
```

```
fit <- lin_pow(dat = galap, compare = TRUE, con = 1)
```

```
summary(fit)
```

```
## Model = Log-log power
##
## Call:
## lm(formula = S ~ A, data = log.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3591 -0.7584  0.1177  0.6009  1.0739
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## LogC   3.01865    0.35442   8.517 6.56e-07 ***
## z       0.33854    0.08523   3.972 0.00139 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7626 on 14 degrees of freedom
## Multiple R-squared:  0.5298, Adjusted R-squared:  0.4962
## F-statistic: 15.78 on 1 and 14 DF, p-value: 0.001391
##
## Power (non-linear) parameters:
## c = 33.18
## z = 0.28
```

```
plot(fit)
```

*#load an example dataset, fit the random placement model and plot the  
#model fit and standard deviation. The 'data' argument requires a  
species-#site abundance matrix: rows are species and columns are  
sites. The area #argument requires a vector of site (island) area  
values.*

```
data(cole_sim)
fit <- coleman(data = cole_sim[[1]], area = cole_sim[[2]])
plot(fit, ModTitle = "Hetfield")
```

*#load an example dataset, fit the GDM using the logarithmic SAR model,  
and #compare the GDM with three alternative (nested) models: area and  
time #(age of each island), area only, and intercept only.*

```
data(galap)
galap$t <- rgamma(16, 5, scale = 2)#add a random time variable
gdm(data = galap, model = "loga", mod_sel = TRUE)
```

```
##
## GDM fit using the logarithmic SAR model
##
## GDM model summary:
##
## Nonlinear regression model
## model: SR ~ Int + A * log(Area) + Ti * Time + Ti2 * Time^2
## data: data
##      Int      A      Ti      Ti2
## -195.933  35.336  38.536  -1.876
## residual sum-of-squares: 73750
##
## Number of iterations to convergence: 1
## Achieved convergence tolerance: 3.146e-08
##
## All model summaries:
##
##           RSE      AIC Delta.AIC
## A           74.44350 187.1908  0.000000
## A + T        77.22723 189.1799  1.989046
## GDM          78.39534 190.3796  3.188759
## Intercept 100.32744 195.8435  8.652648
```

## CONCLUSIONS

The SAR has been a cornerstone of ecological and biogeographical science for almost a century (Arrhenius 1921, MacArthur and Wilson 1967, Rosenzweig 1995) and its form and fit are still of great significance in both theoretical (Storch 2016) and applied



(Chisholm et al. 2018) contexts. The development of the ‘sars’ R package should aid future SAR research by providing a comprehensive set of tools that enable in-depth exploration of SARs and SAR-related patterns. In addition, the package has been designed in such a way as to allow other SAR researchers to add (e.g. via GitHub) new functions and models in the future to ensure the package is of lasting value. For example, future additions to the package could include the suite of countryside biogeography SAR models that have recently been published (Pereira et al. 2014), standard SAR functions not so far incorporated (see Dengler 2009, 2010, Tjørve 2009), or functions specifically intended for analysis of species accumulation curves (Matthews et al. 2016a) or endemics–area relationships. Finally, whilst the focus of this paper has been on classic SARs, there is no reason that the functionality in the ‘sars’ package cannot be used to analyse other diversity–area relationships (e.g. functional or phylogenetic diversity–area relationships, Whittaker et al. 2014, Mazel et al. 2015). Application of the full set of 20 models, in addition to the multimodel SAR framework, included in the ‘sars’ package to a wider range and type of data (e.g. trait and phylogenetic data) will likely be revealing and will help in improving our understanding of SARs, and diversity–area relationships more generally.

#### **DATA ACCESSIBILITY**

The most recent stable version of the package is freely available from CRAN (<https://CRAN.R-project.org/package=sars>), whilst the development version is hosted on GitHub (<https://github.com/txm676/sars>). Version 1.1.1 of the package, presented in this article, has been archived on the Zenodo research data repository (DOI: 10.5281/zenodo.2573067).

## **AUTHORS' CONTRIBUTIONS**

FG and TJM led the design of the study with input on functions required from KAT and RJW. TJM and FG wrote the R package. KAT contributed datasets. TJM wrote the first draft of the paper. All authors contributed to the submitted manuscript.

## **ACKNOWLEDGEMENTS**

Kevin Cazelles, Jürgen Dengler and Michael Borregaard provided comments which greatly improved the package and paper. Jon Sadler kindly tested the R package. The College of Life and Environmental Sciences (University of Birmingham) kindly provided funding for TJM to visit the National and Kapodistrian University of Athens to work on the paper.

## **REFERENCES**

- Arrhenius, O. 1921. Species and area. – *J. Ecol.* 9: 95-99.
- Benchimol, M. and Peres, C.A. 2013. Anthropogenic modulators of species–area relationships in Neotropical primates: a continental-scale analysis of fragmented forest landscapes. – *Divers. Distrib.* 19: 1339-1352.
- Borregaard, M.K., Amorim, I.R., Borges, P.A.V., Cabral, J.S., Fernández-Palacios, J.M., Field, R., Heaney, L.R., Kreft, H., Matthews, T.J., Olesen, J.M., Price, J., Rigal, F., Steinbauer, M.J., Triantis, K.A., Valente, L., Weigelt, P. and Whittaker, R.J. 2017. Oceanic island biogeography through the lens of the general dynamic model: assessment and prospect. – *Biol. Rev.* 92: 830-853.
- Buckland, S.T., Burnham, K.P. and Augustin, N.H. 1997. Model selection: an integral part of inference. – *Biometrics* 53: 603–618.

Burnham, K.P. and Anderson, D.R. 2002. Model selection and multi-model inference: a practical information-theoretic approach, 2nd ed – Springer.

Cardoso, P., Rigal, F. and Carvalho, J.C. 2014. BAT - Biodiversity Assessment Tools, an R package for the measurement and estimation of alpha and beta taxon, phylogenetic and functional diversity. – *Methods Ecol. Evol.* 6: 232-236.

Chisholm, R.A., Lim, F., Yeoh, Y.S., Seah, W.W., Condit, R., Rosindell, J. and He, F. 2018. Species–area relationships and biodiversity loss in fragmented landscapes. – *Ecol. Lett.* 21: 804-813.

Coleman, B.D. 1981. On random placement and species-area relations. – *Mathematical Biosciences* 54: 191-215.

Davison, A.C. and Hinkley, D.V. 1997. Bootstrap methods and their application – Cambridge Univ. Press.

Dengler, J. 2009. Which function describes the species–area relationship best? A review and empirical evaluation. – *J. Biogeogr.* 36: 728-744.

Dengler, J. 2010. Robust methods for detecting a small island effect. – *Divers. Distrib.* 16: 256-266.

Fattorini, S. 2007. To fit or not to fit? A poorly fitting procedure produces inconsistent results when the species-area relationship is used to locate hotspots. – *Biodivers. Conserv.* 16: 2531–2538.

Gitay, H., Roxburgh, S.H. and Wilson, J.B. 1991. Species-area relations in a New Zealand tussock grassland, with implications for nature reserve design and for community structure. – *J. Veg. Sci* 2: 113–118.

Gray, J.S., Ugland, K.I. and Lambshead, J. 2004. On species accumulation and species–area curves. – *Global Ecol. Biogeogr.* 13: 567–568.

Guilhaumon, F., Gimenez, O., Gaston, K.J. and Mouillot, D. 2008. Taxonomic and regional uncertainty in species-area relationships and the identification of richness hotspots. – *Proc. Natl Acad. Sci. USA* 105: 15458–15463.

Guilhaumon, F., Mouillot, D. and Gimenez, O. 2010. mmSAR: an R-package for multimodel species–area relationship inference. – *Ecography* 33: 420–424.

He, F. and Legendre, P. 1996. On species-area relations. – *Am. Nat.* 148: 719–737.

Lomolino, M.V. 2000. Ecology's most general, yet protean 1 pattern: the species- area relationship. – *J. Biogeogr.* 27: 17–26.

MacArthur, R.H. and Wilson, E.O. 1967. The theory of island biogeography – Princeton Univ. Press.

- Matthews, T.J., Cottee-Jones, H.E.W. and Whittaker, R.J. 2015. Quantifying and interpreting nestedness in habitat islands: a synthetic analysis of multiple datasets. – *Divers. Distrib.* 21: 392-404.
- Matthews, T.J., Triantis, K.A., Rigal, F., Borregaard, M.K., Guilhaumon, F. and Whittaker, R.J. 2016a. Island species–area relationships and species accumulation curves are not equivalent: an analysis of habitat island datasets. – *Global Ecol. Biogeogr.* 25: 607-618.
- Matthews, T.J., Guilhaumon, F., Triantis, K.A., Borregaard, M.K. and Whittaker, R.J. 2016b. On the form of species–area relationships in habitat islands and true islands. – *Global Ecol. Biogeogr.* 25: 847–858.
- May, R.M., Lawton, J.H. and Stork, N.E. 1995. Assessing extinction rates – In: J.H. Lawton and R.M. May (eds.), *Extinction rates*. Oxford Univ. Press, pp. 1–24.
- Mazel, F., Renaud, J., Guilhaumon, F., Mouillot, D., Gravel, D. and Thuiller, W. 2015. Mammalian phylogenetic diversity–area relationships at a continental scale. – *Ecology* 96: 2814-2822.
- Niering, W.A. 1963. Terrestrial ecology of Kapingamarangi Atoll, Caroline Islands. – *Ecol. Monogr.* 33: 131-160.

- Pereira, H.M., Ziv, G.U.Y. and Miranda, M. 2014. Countryside species–area relationship as a valid alternative to the matrix-calibrated species–area model. – *Conserv. Biol.* 28: 874-876.
- Preston, F.W. 1962. The canonical distribution of commonness and rarity: Part I. – *Ecology* 43: 185-215.
- R Core Team 2017. R: a language and environment for statistical computing. R foundation for statistical computing: Vienna, Austria.
- Rosenzweig, M.L. 1995. Species diversity in space and time – Cambridge Univ. Press.
- Storch, D. 2016. The theory of the nested species–area relationship: geometric foundations of biodiversity scaling. – *J. Veg. Sci.* 27: 880-891.
- Tjørve, E. 2009. Shapes and functions of species–area curves II: a review of new models and parameterizations. – *J. Biogeogr.* 36: 1435-1445.
- Triantis, K.A., Guilhaumon, F. and Whittaker, R.J. 2012. The island species–area relationship: biology and statistics. – *J. Biogeogr.* 39: 215-231.
- Triantis, K. A., Economo, E. P., Guilhaumon, F., and Ricklefs, R. E. 2015. Diversity regulation at macro-scales: species richness on oceanic archipelagos. – *Global Ecol. Biogeogr.* 24: 594-605.

Wang, Y., Bao, Y., Yu, M., Xu, G. and Ding, P. 2010. Nestedness for different reasons: the distributions of birds, lizards and small mammals on islands of an inundated lake. – *Divers. Distrib.* 16: 862-873.

Whittaker, R. J., Triantis, K. A., and Ladle, R. J. 2008. A general dynamic theory of oceanic island biogeography. – *J. Biogeogr.* 35: 977-994.

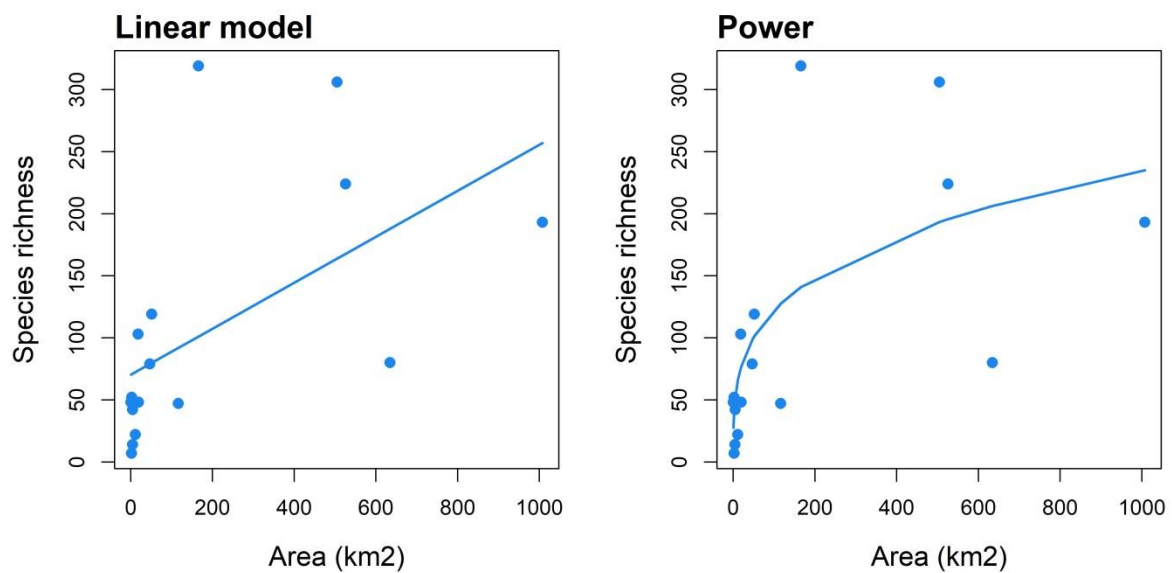
Whittaker, R. J., Rigal, F., Borges, P. A. V., Cardoso, P., Terzopoulou, S., Casanoves, F., . . . Triantis, K. 2014. Functional biogeography of oceanic islands and the scaling of functional diversity in the Azores. – *Proc. Natl Acad. Sci. USA* 111: 13709–13714.

Wickham, H. 2015. *Advanced R* – CRC Press.

Williamson, M., Gaston, K.J. & Lonsdale, W.M. 2001. The species–area relationship does not have an asymptote! – *J. Biogeogr.* 28: 827-830.

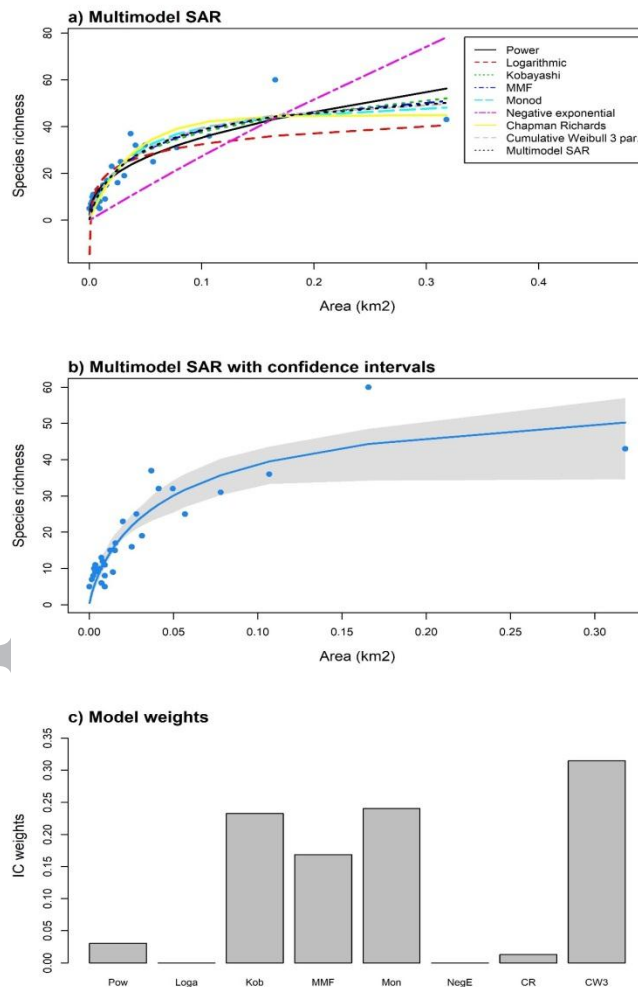
**FIGURE LEGENDS**

**FIGURE 1** The fit (solid lines) of the linear and power SAR models to data of plant species richness (solid circles) on 16 Galapagos islands (Preston 1962; the largest island provided in the source paper has been excluded as it is almost six times larger than the second largest). The code to produce the plots is given as an example in the ‘Fitting individual SAR models and a set of SAR models’ section. The  $R^2$  values for the linear and power models are 0.32 and 0.49.





**FIGURE 2** The different plots that can be generated from an object of class ‘multi’. a) the fit of eight SAR models alongside a multimodel SAR curve generated using  $AIC_c$  weights; b) the multimodel SAR curve plotted on its own with confidence intervals (the confidence intervals are generated using bootstrapping - see details in the main text); c) a barplot of the  $AIC_c$  weights of the different models (the model names on the x-axis match up with legend in (a)). The data are the number of plants species recorded on 32 islands in the Kapingamarangi Atoll (Niering 1963; the shape of the SAR curves differs from those presented in the source paper as Niering used a semi-log plot). The code to produce the plots is given as an example in the ‘Multimodel SAR curve’ section.



## TABLE LEGEND

**TABLE 1** The 20 SAR models available in the ‘sars’ R package. The name of the function to fit each of the models in the package is shown, as is the number of parameters of each model, the model equation, and the model shape. For the model equation, A = sample area, and d, c, z and f are free parameters. Each equation is calculating the number of species. The model shape is the general model shape, as in Triantis et al. (2012). As outlined in the main text, the observed shape can deviate from the general model shape in cases when fitting certain models. This table can be generated in the package using the ‘display\_sars\_models’ function. \*In Triantis et al. (2012), this model was called the exponential model.

Model	Function name	No. parameters	Equation	Model shape
Asymptotic	sar_asymp()	3	$d - c \cdot z^A$	Convex
Beta-P	sar_betap()	4	$d \cdot (1 - (1 + (A/c)^z)^{-f})$	Sigmoid
Chapman–Richards	sar_chapman()	3	$d \cdot (1 - \exp(-z \cdot A)^c)$	Sigmoid
Logarithmic*	sar_loga()	2	$c + z \cdot \log(A)$	Convex
Extended Power 1	sar_epm1()	3	$c \cdot A^{(z \cdot A^d)}$	Convex/Sigmoid
Extended Power 2	sar_epm2()	3	$c \cdot A^{(z - (d/A))}$	Sigmoid
Gompertz	sar_gompertz()	3	$d \cdot \exp(-\exp(-z \cdot (A - c)))$	Sigmoid
Kobayashi	sar_koba()	2	$c \cdot \log(1 + A/z)$	Convex
Linear	sar_linear()	2	$c + z \cdot A$	Linear
Logistic	sar_heleg()	3	$c / (f + A^{(-z)})$	Sigmoid
Monod	sar_monod()	2	$d / (1 + c \cdot A^{(-1)})$	Convex
Morgan–Mercer–Flodin	sar_mmf()	3	$d / (1 + c \cdot A^{(-z)})$	Sigmoid
Negative Exponential	sar_negexpo()	2	$d \cdot (1 - \exp(-z \cdot A))$	Convex
Persistence Function 1	sar_p1()	3	$c \cdot A^z \cdot \exp(-d \cdot A)$	Convex
Persistence Function 2	sar_p2()	3	$c \cdot A^z \cdot \exp(-d/A)$	Sigmoid
Power	sar_power()	2	$c \cdot A^z$	Convex
Power Rosenzweig	sar_powerR()	3	$f + c \cdot A^z$	Convex
Rational	sar_ratio()	3	$(c + z \cdot A) / (1 + d \cdot A)$	Convex
Weibull-3	sar_weibull3()	3	$d \cdot (1 - \exp(-c \cdot A^z))$	Sigmoid
Weibull-4	sar_weibull4()	4	$d \cdot (1 - \exp(-c \cdot A^z))^f$	Sigmoid